



biblio.ugent.be

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

Distributed Coding of Endoscopic Video

Nikos Deligiannis, Frederik Verbist, Joeri Barbarien, Jürgen Slowack, Rik Van de Walle, Peter Schelkens, and Adrian Munteanu

In: Proceedings IEEE International Conference on Image Processing (ICIP), pp. 1853-1856, September 2011.

To refer to or to cite this work, please use the citation to the published version:

Nikos Deligiannis, Frederik Verbist, Joeri Barbarien, Jürgen Slowack, Rik Van de Walle, Peter Schelkens, and Adrian Munteanu (2011). Distributed Coding of Endoscopic Video. Proceedings IEEE International Conference on Image Processing (ICIP), pp. 1853-1856, September 2011.

DISTRIBUTED CODING OF ENDOSCOPIC VIDEO

Nikos Deligiannis,¹ Frederik Verbist,¹ Joeri Barbarien,¹ Jürgen Slowack,²
Rik Van de Walle,² Peter Schelkens,¹ and Adrian Munteanu¹

¹Department of Electronics and Informatics, Vrije Universiteit Brussel – IBBT

²ELIS Department, Multimedia Lab, Ghent University – IBBT

ABSTRACT

Triggered by the challenging prerequisites of wireless capsule endoscopic video technology, this paper presents a novel distributed video coding (DVC) scheme, which employs an original hash-based side-information creation method at the decoder. In contrast to existing DVC schemes, the proposed codec generates high quality side-information at the decoder, even under the strenuous motion conditions encountered in endoscopic video. Performance evaluation using broad endoscopic video material shows that the proposed approach brings notable and consistent compression gains over various state-of-the-art video codecs at the additional benefit of vastly reduced encoding complexity.

Index Terms— Video capsule endoscopy, distributed video coding, hash-driven overlapped block motion estimation.

1. INTRODUCTION

In the last decades, conventional endoscopy, like colonoscopy or gastroscopy, has provided imperative service in the diagnosis and remedy of diverse diseases of the digestive track. Nowadays, remarkable achievements in endoscopic video technology have led to the evolution of wireless capsule endoscopy [1]. A capsule endoscope is a device at the size of a large pill, composed of a limited lifespan battery, a strong light source, an integrated chip video camera, and a radio telemetry transmitter. Once swallowed, the capsule transmits video of the esophagus, stomach and small intestine to a sensor array placed around the patient's abdomen.

The primary clinical use of capsule endoscopy is to examine areas of the small intestine that cannot be visualized by conventional endoscopy. In small intestine pathology, capsule endoscopy has shown superior positive diagnosis rates compared to other methods, such as computed tomographic enteroclysis or magnetic resonance imaging [1]. Furthermore, capsule endoscopy is often preferable in other digestive tract pathologies compared to a more invasive gastroscopy or colonoscopy. In the future, capsule endoscopy is believed to replace conventional endoscopy as a first line investigation for many diseases.

The scale of capsule endoscopic video technology imposes constraints on size and power consumption, calling for coding architectures that deliver low-cost encoding, robustness against transmission errors and high compression efficiency. Current capsule video chips employ conventional coding schemes operating in the low-complexity, intra-frame mode, i.e., Motion JPEG [2]. Anchored in distributed source coding (DSC) fundamentals [3], a strong candidate to fulfill the aforementioned demands is DVC [4], a reverse paradigm enabling separate encoding and joint decoding of video signals.

Endoscopic video content exhibits highly erratic motion characteristics, e.g. low frame acquisition rates, and extreme camera panning caused by gastrointestinal contractions. Most DVC architectures employ motion compensated interpolation (MCI) algorithms to produce side-information at the decoder-side [5], [6], [7]. However, in sequences with irregular motion characteristics, MCI techniques fail to deliver fair prediction quality due to blind motion estimation. To overcome this problem, this work proposes a novel hash-based architecture for distributed coding of capsule endoscopic video. In contrast to existing hash-driven DVC architectures, e.g., [8], [9], the proposed system employs a novel hash-based overlapped block motion estimation (OBME) and compensation algorithm in order to generate multi-hypothesis pixel-based prediction at the decoder. Unlike MCI-based approaches, the proposed technique enables accurate side-information creation even under strenuous conditions. Moreover, in contrast to other hash-driven DVC codecs, e.g., [8], [9], the proposed system does not feature a prediction loop at the encoder, maintaining the error resilience properties of Wyner-Ziv coding. Experimental results on a plethora of endoscopy video sequences show that the proposed codec outperforms the state-of-the-art transform-domain Wyner-Ziv (TDWZ) architecture of [7], as well as Motion JPEG and H.264/AVC Intra, while the required encoding computational resources are radically diminished.

2. DISTRIBUTED SOURCE CODING PRELUDE

DSC frameworks are incited by the information theoretic findings of *Slepian and Wolf* [10] on lossless, separate encoding and joint decoding of correlated sources. *Wyner*

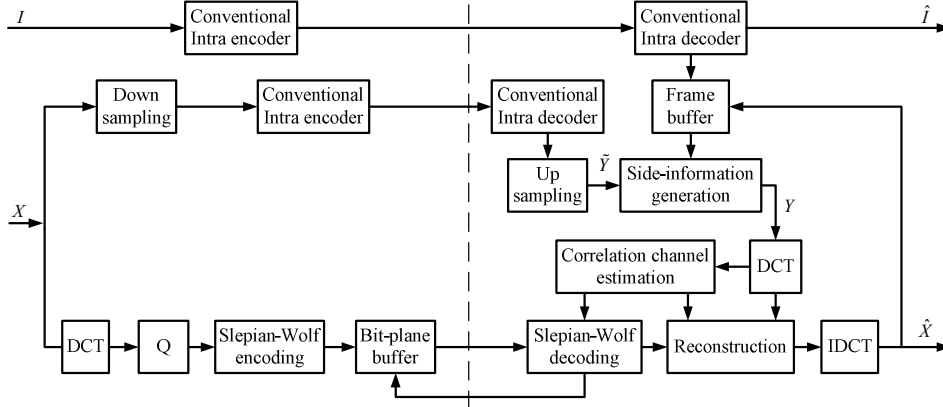


Fig. 1. Schema of the proposed transform-domain hash-based DVC architecture

and Ziv [11] extended these findings to lossy coding with side-information at the decoder. Towards practical Slepian-Wolf (SW) coding, the random binning argument of [10] has been realized using capacity approaching channel codes, such as, Turbo [5] and low-density parity-check (LDPC) [3], [12] codes. Good Wyner-Ziv (WZ) code designs are based on nested quantization followed by an additional layer of binning with channel codes [3]. Intriguingly, [5] has shown that the use of high-rate quantization followed by efficient channel coding leads to a coding loss similar to entropy constraint scalar quantization. Anchored in these findings novel distributed video coding paradigms have been engineered [4], [5].

3. THE PROPOSED DVC CODEC

This paper proposes a novel hash-based transform-domain DVC architecture of which the block diagram is given in Fig. 1. The proposed DVC codec divides the incoming video frames into key and WZ frames. The key frames, I , are conventionally intra coded while the WZ frames, X , are encoded in two parts, i.e., a hash and a WZ bit-stream.

The coded hash information, which is deployed to assist the motion estimation operation at the decoder, consists of a reduced resolution version of each WZ frame coded at a low quality. In particular, in order to form the hash information, the WZ frame first undergoes a down-scaling filter operation with a factor of $d \in \mathbb{Z}_+$. Then, the down-sampled WZ frame is conventionally intra coded at a much lower quality compared to the quality of the key frames.

In addition to the coded hash, a WZ layer is formed for each WZ frame so as to provide compression efficiency, scalability and communication error robustness to the proposed system [3]. In line with Stanford's DVC architecture [5], the WZ frames first undergo a 4×4 integer discrete cosine transform (DCT) and the obtained DCT coefficients are subsequently grouped together into bands. Using a set of predefined quantization matrices (QM), each DCT band, b , is independently quantized with 2^{L_b} levels. A uniform and a double-deadzone scalar quantizer are employed for the DC and the AC bands, respectively. The

quantized symbols are then converted to binary codewords and passed to the SW encoder, which is realized by the LDPC Accumulate (LDPCA) codes of [12].

At the decoder, the key frames are conventionally intra decoded and stored in the reference frame buffer. Thereafter, the decoded and up-scaled hash information is used by the proposed side-information generation module to produce the side-information frame, as detailed in Section 4. Correlation channel estimation is performed online [7]. After SW decoding, the bit-planes of the WZ coded bands are grouped together and optimal minimum mean square error (MMSE) reconstruction [3] followed by inverse DCT are carried out.

4. HASH-DRIVEN OVERLAPPED BLOCK MOTION ESTIMATION AND COMPENSATION

The proposed side-information generation algorithm performs bidirectional OBME using the available hash information and a past and a future reconstructed WZ and/or key frame as reference frames. Similar to other DVC works, e.g., [5], [6], [7], [8], a hierarchical prediction structure is considered. Note that in contrast to our previous work [13], in which motion estimation was bit-plane-based, several changes have been made so as to ensure OBME's compatibility and efficiency with our new DVC approach. These are detailed next.

First, prior to motion estimation, the reference frames are low-pass filtered undergoing the same down- and up-sampling operation with the formation of the hash frame. This is done in order to improve the consistency of the motion vectors, lessening the motion field uncertainty.

To explain the subsequent steps, we introduce the following notations. Let \tilde{Y} be the reconstructed WZ hash, Y be the side-information, and \tilde{R}_k , $k \in \{0,1\}$ be the low-pass filtered versions of the reference frames R_k , respectively. Also, let $Y_{\mathbf{m}}$, $R_{k,\mathbf{m}}$, $\tilde{Y}_{\mathbf{m}}$, $\tilde{R}_{k,\mathbf{m}}$ be blocks of size $B \times B$ pixels with top-left coordinates \mathbf{m} in Y , R_k , \tilde{Y} and \tilde{R}_k , respectively. Also, denote by $Y_{\mathbf{m}}(\mathbf{s})$ the sample at position $\mathbf{s} = (i, j)$ in the block $Y_{\mathbf{m}}$.

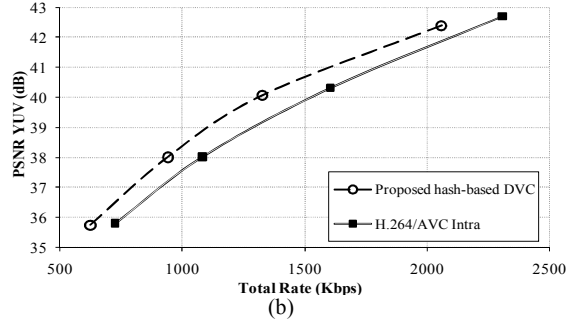
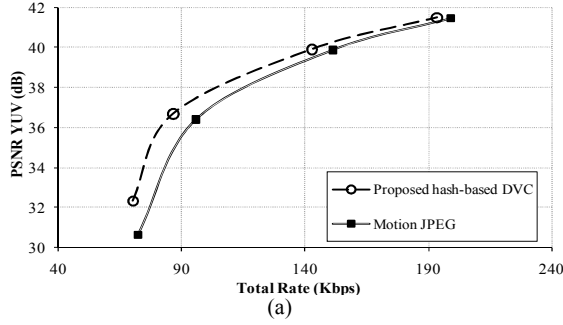


Fig. 2. Compression performance of the proposed DVC codec in (a) capsule endoscopic video material (BD gains: 9.33%, 0.97dB), and (b) conventional endoscopic (BD gains: 12.41%, 0.77dB); $RSNR_{YUV} = (4 \times PSNR_Y + PSNR_U + PSNR_V) / 6$.

Using a specified overlapping step $\varepsilon \in \mathbb{Z}_+$, $1 \leq \varepsilon \leq B$, the available hash frame is divided into overlapping spatial blocks, \tilde{Y}_u , with top-left coordinates $\mathbf{u} = (u_1, u_2)$. For each overlapping block \tilde{Y}_u , the best matching block within a search range ρ , is found in the filtered reference frames \tilde{R}_k . Contrary to [13], the employed matching criterion minimizes the sum of absolute differences (SAD) between \tilde{Y}_u and a block $\tilde{R}_{k, \mathbf{u}-\mathbf{v}}$, $\mathbf{v} = (v_1, v_2)$, $-\rho < v_1, v_2 \leq \rho$ in the filtered version of the reference frames. The ensuing motion vector \mathbf{v} is

$$\mathbf{v} = \arg \min_{\mathbf{v}} \sum_{\mathbf{s}} |\tilde{Y}_u(\mathbf{s}) - \tilde{R}_{k, \mathbf{u}-\mathbf{v}}(\mathbf{s})|. \quad (1)$$

The derived integer-pel motion field is then extrapolated to the unfiltered reference frames, and used to derive the best temporal predictor, for each overlapping block Y_u , of the side-information. After the execution of the OBME, each pixel $Y(\mathbf{s})$, $\mathbf{s} = (i, j)$ in the side-information frame belongs to a number of overlapping blocks Y_{u_n} with $\mathbf{u}_n = (u_{n,1}, u_{n,2})$. For each of these blocks, OBME has identified a temporal predictor block R_{k, \mathbf{u}_n} in one reference frame. This means that, each pixel $Y(\mathbf{s})$ in the side-information has a number of corresponding predictors r_{k, \mathbf{u}_n} in the blocks R_{k, \mathbf{u}_n} .

In the compensation process, one combines the derived predictors per pixel to perform multi-hypothesis pixel-based prediction. In particular, each pixel $Y(\mathbf{s})$ in the side-information frame is determined as the mean value of the predictor values r_{k, \mathbf{u}_n} as:

$$Y(\mathbf{s}) = \frac{1}{N_{k, \mathbf{u}_c}} \sum_{\mathbf{u}_n} r_{k, \mathbf{u}_n}, \quad (2)$$

where, N_{k, \mathbf{u}_c} denotes the number of predictors for the side-information pixel $Y(\mathbf{s})$. Note that the derived multi-hypothesis motion field is also employed to estimate the side-information frame's chroma components from the chroma components of the reference frames.

Compared to existing works, e.g., [5], [7], [8], [9], the proposed bidirectional OBME technique mitigates the motion uncertainty at a pixel level by not limiting the motion information to just one motion vector for an entire block. Moreover, blocking artifacts are drastically reduced, increasing the subjective quality, which is a strong prerequisite in medical imaging applications.

Table I. Average BD deltas of the proposed DVC codec against TDWZ [7] and H.264/AVC Intra (only the luma component is coded).

Sequence	vs. TDWZ [7]		vs. H.264/AVC	
	ΔR (%)	$\Delta PSNR$ (dB)	ΔR (%)	$\Delta PSNR$ (dB)
Endoscopic video 1	-43.3	3.56	-4.1	0.29
Endoscopic video 2	-43.13	3.58	2.05	-0.15

5. EXPERIMENTAL RESULTS

This section assesses the coding efficiency of the proposed DVC codec against several state-of-the-art low-cost encoding video codecs, namely, Motion JPEG, the TDWZ codec of [7], and H.264/AVC Intra [14]. In the proposed DVC codec, the QMs of the WZ frames and the quantization parameters of the key frames are meticulously selected so as to retain high and quasi-constant decoded frames' quality, as required by medical applications. Moreover, a GOP size of 2, an overlapping step of $\varepsilon = 8$, a search range of $\rho = 20$ and a block size of $B = 32$ pixels are selected.

To appraise its applicability in wireless capsule endoscopy, the proposed codec has been evaluated using extensive capsule endoscopic video material visualizing diverse areas of the small bowel. The capsule acquisition rate was 5Hz with a frame resolution of 256×256 pixels. Motion JPEG has been set as benchmark, since this codec is used in current capsule endoscopes [2]. For a fair comparison, Motion JPEG has also been used to code the key frames and the WZ hash in the proposed codec. The results shown in Fig. 2(a), which are consistent across different sequences, point out that the proposed DVC yields notable Bjøntegaard (BD) [15] improvements of 9.33% and 0.97dB against Motion JPEG.

New trends in capsule endoscopy aim at increasing the frame rates and resolutions. Therefore, to explore its capability under these conditions, the proposed DVC codec is evaluated using conventional endoscopic video sequences monitoring diverse parts of digestive tracks of several patients. The obtained sequences exhibit a frame rate of 30Hz and a resolution of 480×320 pixels. In this experiment, the proposed codec employs H.264/AVC Intra to code the key frames and the hash information is coded with $QP_{\text{hash}} = QP_{\text{intra}} + 3$. The experimental results – see Table I,

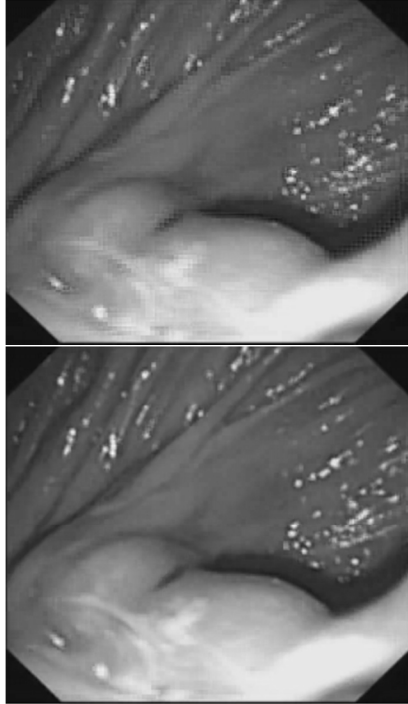


Fig. 3. Visual performance for endoscopic video coded at 1250kbps with; (top) TDWZ [7], 35.4dB, (bottom) proposed DVC, 39.1dB.

show that the proposed codec delivers impressive average BD rate and PSNR [15] gains against our implementation of the TDWZ codec of [7]. We note that the codec of [7] provides state-of-the-art MCI-based DVC performance, being the basis of the DISCOVER codec [6]. We note also that DISCOVER could not be included in the comparison since it does not support the frame resolution of the video data. Compared to H.264/AVC Intra, the experimental results – see Table I and Fig. 2(b), show that the proposed codec delivers average rate and PSNR improvements of up to 12.41% and 0.77dB, respectively.

A visual comparison between TDWZ [7] and the proposed codec for a WZ frame is given in Fig. 3. The proposed codec yields significantly better visual quality and does not suffer from blocking artifacts, typically affecting TDWZ [7] at this rate.

Complexity tests confirm that the proposed codec offers a notable reduction of the encoding computational time, i.e., approximately 40%, compared to H.264/AVC Intra. When compared to Motion JPEG, the proposed codec, although currently not optimized for speed, exhibits similar encoding computational time.

6. CONCLUSIONS

This work addresses the potential of DSC principles in wireless capsule endoscopy. A novel DVC design, which encodes a robust hash to assist side-information generation at the decoder, is presented. Counter to existing works, the proposed system embraces a powerful side-information creation method which performs overlapped block motion

estimation and compensation. Detailed experimentation shows that the proposed codec outperforms various state-of-the-art codecs, while maintaining low encoding complexity.

7. REFERENCES

- [1] R. Sidhu, D. Sanders, and M. McAlindon, "Gastrointestinal capsule endoscopy: from tertiary centres to primary care," *BMJ*, pp. 528-531, Mar. 2006.
- [2] D. Turgis and R. Puers, "Image compression in video radio transmission for capsule endoscopy," *Sensors and Actuators A: Physical*, pp. 129-136, Sep. 2005.
- [3] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, pp. 80-94, Sep. 2004.
- [4] F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi, and S. Klomp, "Distributed video coding: selecting the most promising application scenarios," *Signal Processing: Image Communication*, vol. 23, pp. 339-352, Jun. 2008.
- [5] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, pp. 71-83, Jan. 2005.
- [6] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Quaret, "The DISCOVER codec: Architecture, techniques and evaluation," in *Picture Coding Symposium*, Lisboa, Portugal, 2007.
- [7] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 1177-1190, Sep. 2008.
- [8] J. Ascenso, C. Brites, and F. Pereira, "A flexible side information generation framework for distributed video coding," *Springer, Multimedia Tools and Applications, Special Issue on Distributed Video Coding*, vol. 48, pp. 381-409, Jul. 2010.
- [9] B. Macchiavello, D. Mukherjee, and R. L. de Queiroz, "Iterative side-information generation in a mixed resolution Wyner-Ziv framework," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, pp. 1409-1423, Oct. 2009.
- [10] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471-480, Jul. 1973.
- [11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1-10, Jan. 1976.
- [12] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Processing*, vol. 86, pp. 3123-3130, 2006.
- [13] N. Deligiannis, A. Munteanu, T. Clerckx, J. Cornelis, and P. Schelkens, "Overlapped block motion estimation and probabilistic compensation with application in distributed video coding," *IEEE Signal Processing Letters*, vol. 16, pp. 743-746, Sep. 2009.
- [14] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560-576, Jul. 2003.
- [15] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T Video Coding Experts Group (VCEG), Austin, TX, Document VCEG-M33, Apr. 2001.